



Institut Agronomique
Néo-Calédonien

Session de formation aux méthodologies statistiques
Institut Agronomique Calédonien, Païta, Nouvelle-Calédonie

du 14 au 18 avril 2003

Cécile DUBOIS et Samir MESSAD

Rapport CIRAD-EMVT n°2003-024

juin 2003



CIRAD-EMVT (TA 30/A)
CIRAD-FLHOR (TA 50/PS4)
34398 Montpellier Cedex 5
France

AUTEUR(S) : Cécile DUBOIS
Samir MESSAD

ACCES AU DOCUMENT :
au service de documentation du CIRAD-FLHOR et
CIRAD-EMVT

ORGANISME AUTEUR : CIRAD-EMVT
CIRAD-FLHOR

ACCES A LA REFERENCE DU DOCUMENT : Libre

ETUDE FINANCEE PAR : CIRAD-Direction

REFERENCE : OM n° 30 06 03 102 (S. Messad) et OM n° 50 93 03 186 (C. Dubois)

AU PROFIT DE : IAC (Institut Agronomique Néo-Calédonien).

TITRE : Session de formation aux méthodologies statistiques – Rapport de mission Cirad-Emvt
n°2003-024

TYPE D'APPROCHE : Mission de formation.

DATE ET LIEU DE PUBLICATION : juin 2003, Montpellier (France).

PAYS OU REGIONS CONCERNÉS : Nouvelle-Calédonie.

MOTS CLES : Formation, Statistique, Biométrie, Suivi d'élevages, Essais agronomiques, Enquêtes

RESUME :

La session de formation « méthodologies statistiques » préparée et dispensée par Cécile Dubois, statisticienne au CIRAD-Flhor et Samir Messad, statisticien au CIRAD-EMVT, s'est déroulée sur cinq journées. L'objectif principal était de faire acquérir une vision plus large de l'utilisation, et de l'utilité, de la biométrie pour répondre à des questions zootechniques ou agronomiques. Elle s'est appuyée sur deux logiciels de traitement statistique : R (langage et environnement de calcul statistique et graphique) et Minitab.

L'intervention conjointe de statisticiens de deux départements différents du Cirad a permis de répondre aux attentes des chercheurs et techniciens des différents programmes de l'Institut Agronomique Néo-Calédonien (IAC) et malgré la relative disparités des niveaux de bases, des thématiques et du nombre de participants, il nous a semblé que le message d'une démarche statistique modulable selon les objectifs et les données, et non universellement pré-établie, a bien été perçu par l'ensemble des stagiaires. D'une manière plus générale, cette formation était donc l'occasion d'un échange qui devrait permettre pour l'avenir de mieux comprendre les questions et attentes des chercheurs et des biométriciens.

Sommaire

Remerciements.....	1
Personnes rencontrées.....	1
Déroulement de la mission conjointe S. Messad (SM), C. Dubois (CD)	2
Introduction.....	4
Liste des participants	5
Déroulement de la session de formation aux méthodologies statistiques.....	6
Outils de base de la statistique. Les résumés graphiques et numériques.....	6
Premiers contacts avec le logiciel de statistique R.....	6
Introduction aux tests d'hypothèses : étude des tableaux 2x2	7
Introduction aux plans d'expérience et ANOVA	7
Initiation à l'analyse des données	7
Evaluation.....	8
Conclusion/Recommandations	10
Bibliographie	11
Annexes.....	12
Annexe 1. Document d'évaluation de la formation reçue.	13
Annexe 2. Sommaire des fiches de formation.....	14

Remerciements

Nous tenons à remercier l'ensemble des personnes qui nous ont accueillis et fait en sorte que notre séjour se déroule dans de très bonnes conditions. Remerciements particuliers à Thierry MENESSION, directeur de l'IAC, Nicolas BARRE, chef du programme Elevage de l'IAC, François MADEMBA-SY, Chargé de mission du Cirad en Nouvelle-Calédonie, et bien sûr les participants pour leur écoute attentive et intéressée.

Personnes rencontrées

Thierry MENESSION, directeur de l'IAC, Nouvelle-Calédonie.

Nicolas BARRE, chef du programme Elevage de l'IAC, Nouvelle-Calédonie.

François MADEMBA-SY, chargé de mission du Cirad en Nouvelle-Calédonie, chef du programme Cultures Fruitières de l'IAC, Nouvelle-Calédonie.

Déroulement de la mission conjointe S. Messad (SM), C. Dubois (CD)

30/03/2003	Départ de Montpellier de l'intervenant EMVT : Samir MESSAD
31/03/2003	Arrivée à Nouméa (Nouvelle-Calédonie) Rencontre avec Thierry MENESSON, directeur de l'IAC et Nicolas BARRE, Chef du programme Elevage à l'IAC et coordinateur de l'organisation de la formation Mission d'appui méthodologique à l'équipe Elevage en collaboration avec Jean-Paul POIVEY sur l'enquête suivi résistance des bovins aux tiques en Nouvelle-Calédonie (OM n°300603101). (SM)
07/04/2003	Départ de Montpellier de l'intervenant FLHOR : Cécile Dubois
09/04/2003	Arrivée à Nouméa. (CD)
09 et 10/04/2003	Station de Recherches Fruitières de Pocquereux : rencontre avec les chercheurs et techniciens de la station (F. Mademba-Sy, S. Lebégin, C. Mille, V. Kagy, M. Bouteiller, Z. Lemerre-Desprez). Visite du domaine et de parcelles expérimentales, présentation des thématiques biologiques étudiées, problèmes statistiques rencontrés. (CD)
11/04/2003	Station de Recherches du Mont Dore : rencontre avec les chercheurs du programme Cultures Maraîchères et Horticoles (L. Desvals, Gildas Gateblé, Marc Deschamps). Visite des installations (parcelles expérimentales sous abris), présentation des thématiques biologiques étudiées, problèmes statistiques rencontrés. (CD)
14/04/2003	Début de la formation Introduction de Nicolas BARRE Présentation détaillée du programme et discussion avec les participants. - Outils de base de la statistique. Les résumés graphiques et numériques. - Premiers contacts avec le logiciel de statistique R
15/04/2003	- Introduction aux tests d'hypothèses : étude des tableaux 2x2
16/04/2003	- Introduction aux plans d'expérience et ANOVA
17 et 18/04/2003	- Initiation à l'analyse des données - Arbres de régression et arbres de classification
18/04/2003	Bilan de la formation avec les participants (évaluation) et Nicolas BARRE.

19/04/2003	Départ de S. Messad de Nouméa
20/04/2003	Arrivée à Montpellier de S. Messad
22 et 24/04/2003	Station de Recherches Fruitières de Pocquereux : application pratique de méthodes vues pendant la formation sur des données expérimentales. (CD)
28/04/2003	Départ de C. Dubois de Nouméa
29/04/2003	Arrivée à Montpellier de C. Dubois

Introduction

Cette session de formation vient en réponse à la demande exprimée par les chercheurs de l'IAC en matière de formation aux méthodologies statistiques. La nature de cette demande était cependant difficile à cerner du fait de l'hétérogénéité des thématiques, des besoins méthodologiques et des niveaux des participants. On connaît la difficulté de répondre par de la formation à des besoins très différents : les méthodes et les démarches statistiques même si elles sont souvent présentées dans la littérature de façon standard, le sont rarement dans les faits.

Pour répondre le plus efficacement possible, il a été décidé l'intervention de deux formateurs-statisticiens dont les départements d'origine recouvrent les profils des chercheurs et techniciens qui ont participé à la formation. Cécile Dubois, statisticienne au CIRAD-FLHOR et Samir Messad, statisticien au CIRAD-EMVT ont assuré cette session de formation en abordant un large spectre de méthodes statistiques et en tentant dans la mesure du possible de présenter des démarches qui appartiennent aux disciplines d'applications : l'agronomie et l'amélioration des systèmes de cultures pour la partie « végétale », la zootechnie et l'écopathologie pour la partie « animale ».

Le programme s'est appuyé sur le logiciel **R** (langage et environnement de calcul statistique et graphique) et le logiciel **MINITAB** © ainsi qu'un ensemble de fiches de travaux dirigés et de travaux pratiques réalisés sur des jeux de données fictifs et réels.

Ce document présente les contenus et les appréciations d'ensemble des formateurs.

Liste des participants

NOM	DISCIPLINE	ORGANISME	LOCALISATION
BOUTEILLER Marcelle	Technicienne Qualité Fruits (post-récolte)	IAC	Pocquereux
BRESCIA Fabrice	Zoologie/Ecologie	IAC	Port-Laguerre
CORNU Alain	Agronomie	IAC	Maré
de GARINE WICHATITSKY Michel	Ecologie animale	CIRAD	Port-Laguerre
DESMOULINS Frédéric	Ecologie Cerf rusa	IAC	Port-Laguerre
DESVALS Laurent		IAC	Mont Dore
GATEBLE Gildas	Horticulture ornementale	IAC	Mont Dore
KAGY Valérie	Post-récolte + phytopathologie	IAC	Pocquereux
KOJFER Laurent	Agronomie	AICA	Bourail
LEBEGIN Stéphane	Tech. Agrophysiologie	IAC	Pocquereux
MADEMBA-SY François	Agrophysiologie	CIRAD	Pocquereux
MAGRON Franck	Informatique/Halieutique	CPS	Nouméa
MILLE Christian	Entomologiste	IAC	Pocquereux
SPAGGIARI Jérôme	Ecologie Cerf rusa	IAC	Port-Laguerre

CPS = Commission du Pacifique

AICA = Association Inter-provinciale de gestion des Centres Agricoles

IAC = Institut Agronomique néo-Calédonien

Déroulement de la session de formation aux méthodologies statistiques

La formation s'est déroulée sur 5 jours et s'est appuyée sur le logiciel **R** qui est à la fois un langage et environnement de calcul statistique et graphique. Certaines applications qui concernent les méthodes d'analyse de variance ont été traitées avec le logiciel de statistique MINITAB ©. Des fiches de travaux dirigés, un document d'initiation à l'analyse des données et des jeux de données fictifs et réels ont servi de support aux présentations.

Ce cours s'appuie notamment sur des fiches qui se veulent indépendantes et qui abordent différents aspects du travail de traitement des données. Il peut s'agir de l'initiation à une méthodologie ou un groupe de méthodes statistiques ou à des aspects plus opérationnels comme la prise en main du logiciel d'analyse des données **R**.

Ainsi, c'est sur ces documents que se fonde notre approche pédagogique qui ne se contente pas d'un cours de statistique mais qui tente également d'être un transfert de nos connaissances et notre approche en matière de traitement des données.

Outils de base de la statistique. Les résumés graphiques et numériques.

Une petite présentation introductive tirée de l'ouvrage de Claudine Robert : L'empereur et la girafe, Leçons élémentaires de statistiques, Diderot Edition, propose une révision ludique des notions de base en statistique comme les caractères de centralité (moyenne et médiane), de dispersion (variance) et du bon usage des principaux outils statistiques graphiques.

Cette partie a permis d'entamer le dialogue nécessaire entre biométriciens et thématiciens. Une discussion a permis de différencier les types d'approche en analyse statistique. Ainsi l'approche exploratoire ou descriptive et l'approche inférentielle (modélisation) qui sont souvent distinguées ont été exposées au travers de leurs différences mais également de leurs complémentarités. Nous avons particulièrement insisté sur les fondements d'une approche adaptative : il y a souvent plusieurs méthodes pour traiter un problème et le choix des méthodes est toujours guidé par les objectifs, c'est-à-dire des hypothèses clairement énoncées en amont du travail d'analyse.

Premiers contacts avec le logiciel de statistique R

Sous forme de travaux dirigés, les participants ont fait leurs premières armes avec le logiciel **R**. Ce logiciel offre un cadre très complet pour le spécialiste, mais il permet également aux utilisateurs, à condition d'un investissement pour acquérir une autonomie minimale, une grande souplesse d'utilisation. Il faut également dire que le choix de ce logiciel s'est fait pour des raisons pédagogiques, l'enseignement universitaire de la statistique est souvent réalisé avec des logiciels de ce type (voir également SAS © ou Splus ©). Cette facilité pédagogique trouve également sa raison d'être dans la possibilité d'échange de procédures de traitements adaptées.

- Installer **R** sur son PC,
- Première session de travail avec **R**,
- Importer et exporter des données dans **R**.

Introduction aux tests d'hypothèses : étude des tableaux 2x2

Au travers de deux exemples d'étude des tableaux 2x2, cette partie propose de mieux comprendre la construction de différents types d'indicateurs (probabilité conditionnelle, risque relatif et odds ratio) et de l'application d'un test statistique (notamment au travers du test du khi deux). Cette fiche est à la fois une application pratique d'analyse d'un type de tableau que l'on trouve fréquemment en épidémiologie comme un tableau de fréquence allélique pour la détection de gène marqueur d'une maladie et un tableau d'incidence de la mortalité issu d'un suivi d'une cohorte d'animaux.

Introduction aux plans d'expérience et ANOVA

Présentation et travaux dirigés autour des méthodologies des plans d'expérience et des techniques de traitement par l'analyse de variance. Cadre théorique et hypothèses, applications et recommandations. Cette présentation se limite au cas des dispositifs expérimentaux les plus fréquemment utilisés : randomisation totale et blocs randomisés, avec 1 facteur ou 2 facteurs étudiés.

Initiation à l'analyse des données

Parmi les méthodes descriptives nous avons insisté sur l'intérêt des méthodes de l'analyse des données pour la prise en compte des phénomènes multivariés, la construction de typologies et leur pouvoir synthétique lorsque l'on dispose d'une information en quantité importante comme par exemple les fichiers de données d'enquête.

Il s'agissait donc d'un premier aperçu des méthodes de base d'analyses factorielles, de la classification automatique et d'arbres de décision :

- Introduction aux méthodes factorielles : l'Analyse en Composantes Principales (ACP) et l'Analyse Factorielle des Correspondances (AFC),
- Méthode de synthèse des questionnaires d'enquêtes typologiques : l'Analyse Factorielle des Correspondances Multiples (AFCM).
- Méthode de classification automatique : la Classification Ascendante Hiérarchique (CAH).
- Arbres de régression, arbres de classification.

Evaluation

Afin d'avoir une idée de la perception du cours par les participants, nous avons procédé à une petite évaluation qui permet non seulement d'identifier les lacunes et les éléments qui ont été compris dans le cours mais également à tenter de cerner l'attitude et les motivations des stagiaires sur leur travail d'apprentissage en matière de méthodologies.

Expérience personnelle dans le traitement statistique avant ce stage :

1. <i>quasi-inexistante</i>	0 réponse
2. <i>occasionnelle</i>	8 réponses
3. <i>régulièrement</i>	2 réponses
4. <i>systématique dans tous vos travaux</i>	1 réponse

La plupart du temps, la demande de méthodologies pour gérer et procéder à des traitements statistiques est occasionnelle de la part des chercheurs. Ces aspects sont difficilement contournables et c'est précisément ce qui explique la demande. Nous nous sommes efforcés d'expliquer quel investissement personnel minimal nous semblait nécessaire pour l'acquisition des méthodes essentielles et comment pouvait fonctionner la collaboration entre chercheur et statisticien.

Qu'avez-vous apprécié au cours de ce stage ?

L'appréciation du logiciel de statistique **R** et de ses principes de fonctionnement nous conforte dans notre choix d'un outil puissant, évolutif et libre de tous droits. Les difficultés apparentes d'apprentissage du logiciel sont presque balayées. Presque, car pendant la session, des digressions ont permis d'appréhender le niveau de complexité du langage.

Les participants ont largement apprécié la présentation d'un éventail assez large de méthodologies, ce qui était un objectif avoué et semblait dès le départ être une façon de répondre à une demande multiple en matière de méthodes.

Enfin, les travaux pratiques sur PC ont été également appréciés. Il semble donc que l'équilibre théorique-pratique ait convenu à la plupart.

Qu'avez-vous peu apprécié au cours de ce stage ?

A la fois la densité du programme de la formation et celle des travaux dirigés. Certains évoquent ainsi un manque de recul et de mise à l'épreuve devant un problème statistique concret. Peut être aurait-il fallu proposer des travaux pratiques mais l'exercice était difficile car il aurait fallu pouvoir travailler avec un comité plus restreint et se donner suffisamment de temps afin que la partie théorique et travaux dirigés soient au préalable « digérés ».

Les problèmes d'interprétation statistique-interprétation biologique. Nous pensons que l'objectif final est précisément de pouvoir répondre à cette problématique. C'est un discours conjoint et commun du statisticien et du biologiste qui permettra d'avancer sur cette question.

Qu'est-ce qui vous semblait difficile dans le contenu de ce stage ?

Le langage du logiciel **R**, le vocabulaire parfois même qualifié de « jargon de spécialiste » ont rendu, selon les participants, ce cours assez difficile. Il n'est bien évidemment pas question de former des spécialistes mais cette difficulté est contrebalancée par de nombreux avantages qui nous semblent déterminants dans les échanges entre techniciens, chercheurs et statisticiens. L'enjeu de cette formation était également d'initier une autre façon de travailler et d'appréhender le travail d'analyse des données. Moins comme une discipline qui doit rendre compte de manière standardisée d'un phénomène biologique : une décision à prendre à partir d'un calcul fait sur les données, mais plus comme un outil supplémentaire de la recherche pour faire état ou comprendre des faits biologiques.

Qu'est-ce qui vous semblait facile dans le contenu de ce stage ?

Peu de réponses à cette question, sauf celle jugeant le niveau théorique proposé en-deçà de celui souhaité. Malheureusement, les niveaux sont toujours hétérogènes dans ce genre de formation et nous avons pris l'option d'un cours d'initiation. A condition d'un niveau de base minimum, les méthodologies plus avancées peuvent s'acquérir seul ou dans le cadre d'un travail étroit avec un biométricien, investissement personnel sur lequel nous avons beaucoup insisté lors de cette session.

Face à des données zootechniques ou épidémiologiques et des méthodes d'analyses présentées au cours de ce stage, vous vous sentez ...

- | | |
|---|------------|
| 1. pas du tout à l'aise. | 0 réponse |
| 2. pas trop à l'aise. | 1 réponse |
| 3. un peu plus à l'aise. | 3 réponses |
| 4. avoir bien compris l'ensemble des méthodes exposées sans savoir toutefois si vous saurez les mettre en œuvre seul. | 5 réponses |
| 5. avoir acquis des techniques et une démarche. | 2 réponses |

La réponse n°4 correspond d'une certaine manière aux ambitions de cette formation. Acquérir un minimum d'autonomie en méthodologie statistique commence par avoir une vue plus claire et sereine sur cette discipline qui est aux yeux de certains un mal nécessaire.

Comment envisageriez-vous une suite à ce stage ?

On peut citer, les réponses suivantes :

- Tenter d'éprouver les méthodes et les logiciels sur les données dont disposent les stagiaires,
- Solliciter les intervenants lors des phases de questionnements sur l'application des méthodes sur leur propre jeu de données,
- Envisager une seconde session de formation « avancée »,
- Recruter un(e) statisticien(ne) à l'IAC.

Conclusion/Recommandations

Nous faisons état dans cette conclusion de quelques remarques, dont certaines déjà citées pour d'autres missions de formation, quant à notre perception de la formation des chercheurs aux outils statistiques.

Améliorer la compréhension entre « biologistes » et statisticiens

Somme toute, avoir proposé un cours à un public hétérogène quant aux disciplines a été en tous points positif. L'adaptation d'une méthode à des contextes thématiques différents peut s'avérer tout à fait enrichissante et certains participants l'ont fait remarquer. Même si les stagiaires ont appréhendé cette double intervention comme un cours de statistique et donc n'ont pas eu une participation aussi active que souhaitée, il semble que chacun a pu se forger une vision plus juste et plus réaliste du travail de traitement statistique.

Méthodologie statistique versus logiciel statistique

Contrairement à une idée répandue chez les utilisateurs, l'aspect logiciel n'est pas fondamental. Une approche statistique consiste d'abord en des choix méthodologiques, choix conditionnés à de multiples critères rendant parfois complexe le traitement statistique. Parce qu'ils tentent d'être exhaustifs dans les résultats proposés, les utilisateurs ont tendance à penser que les logiciels statistiques et donc leur apprentissage leur permettront de répondre à ces questions. En réalité, les logiciels proposent, à travers la documentation qui les accompagne, rarement des démarches ou des exposés détaillés sur les méthodes. L'apprentissage d'un logiciel n'est pas suffisant. Il était donc bien question d'un apprentissage de concepts, méthodes, démarches et approches autour de la statistique et non celui d'un logiciel qui est somme toute une boîte à outils plus ou moins complète.

Acquisition d'un minimum d'autonomie

S'agissant des formations en général et celle en analyse des données en particulier, il faut insister sur le fait que les méthodologies présentées au cours de ces sessions doivent faire l'objet d'une appropriation par les participants. Les « recettes » en ce domaine n'existent pas et nous nous sommes attachés à présenter des méthodes mais également une démarche méthodologique qui est certes celle du statisticien mais qui permet de mieux comprendre comment l'outil statistique peut être mis en œuvre. C'est une première étape pour les participants qui doivent faire l'effort de transposition, en espérant qu'il sera plus aisé après avoir beaucoup insisté sur les aspects opérationnels de ces méthodes.

Bibliographie

Paradis E., 2002. R pour les débutants. Document distribué par Internet. URL: http://cran.r-project.org/doc/contrib/Rdebuts_fr.pdf

Robert C., 1998. L'empereur et la girafe, Leçons élémentaires de statistiques, Diderot Edition.

Ross Ihaka and Robert Gentleman, 1996. R: A language for data analysis and graphics. Journal of Computational and Graphical Statistics, 5(3): 299-314.

Site web de présentation du projet **R** : <http://www.r-project.org/>

Site web de téléchargement du logiciel **R** : <http://cran.r-project.org/>

Annexes

Annexe 1. Document d'évaluation de la formation reçue.

Session de formation aux méthodologies statistiques

14-18 avril 2003

Institut Agronomique Calédonien (IAC), Nouvelle-Calédonie

Expérience personnelle dans le traitement statistique avant ce stage :

1. quasi-inexistante 2. occasionnelle 3. régulièrement 4. systématique dans tous vos travaux

Qu'avez-vous apprécié au cours de ce stage ?

Qu'avez-vous peu apprécié au cours de ce stage ?

Qu'est-ce qui vous semblait difficile dans le contenu de ce stage ?

Qu'est-ce qui vous semblait facile dans le contenu de ce stage ?

Face à des données zootechniques ou épidémiologiques et des méthodes d'analyses présentées au cours de ce stage, vous vous sentez ...

- pas du tout à l'aise.
- pas trop à l'aise.
- un peu plus à l'aise.
- avoir bien compris l'ensemble des méthodes exposées sans savoir toutefois si vous saurez les mettre en œuvre seul.
- avoir acquis des techniques et une démarche.

Comment envisageriez-vous une suite à ce stage ?

Annexe 2. Sommaire des fiches de formation

Installer R sur son PC

1. Objectif de la fiche
2. Qu'est ce que R ? (Paradis E., 2000)
3. Installation du logiciel R (Chessel D., 2001)
4. Personnaliser son dossier de travail
5. Installation d'une librairie
6. Configurer R
7. Désinstaller R de son PC
8. Bibliographie

Première session de travail avec R

1. Objectif de la fiche
2. Consulter la documentation
3. Utiliser l'historique des commandes
4. Outils basiques de R
 - 4.1. L'opérateur <-
 - 4.2. Le collecteur c
 - 4.3. Lister et effacer des objets R
 - 4.4. Les objets R
5. Création et manipulation des tableaux de données sous R
 - 5.1. Importer des données dans R
 - 5.2. Accéder aux données du data frame
 - 5.3. Les variables qualitatives ou facteurs dans un data frame
 - 5.4. Modifier les noms des lignes ou des colonnes d'un data frame
 - 5.5. Exporter des données
6. Ecrire une fonction personnelle
7. Bibliographie

Importer et exporter des données dans R

1. Objectif de la fiche
2. Importation des fichiers textes
 - 2.1. Préparation des fichiers textes
 - 2.2. La fonction read.table
 - 2.3. Construction des tableaux de contingence
3. Importation des données dans R depuis une source ODBC
 - 3.1. Définir une base source ODBC
 - 3.2. La librairie RODBC
4. Exportation des données depuis R
5. Bibliographie

Etude des tableaux de contingence 2 x 2

1. Introduction
2. Cas des grands échantillons : étude de l'effet de la vermifugation sur la mortalité naturelle
 - 2.1. Données et notations
 - 2.2. Indicateurs et tests statistiques
 - 2.2.1. Notion d'indépendance entre deux variables qualitatives
 - 2.2.2. Le nombre de degrés de liberté
 - 2.2.3. Le test du khi deux
 - 2.2.4. Probabilités conditionnelles
 - 2.2.5. Le risque relatif et l'odds ratio
3. Cas des petits échantillons : identification d'un marqueur génétique de résistance à la dermatophilose chez les bovins
 - 3.1. Test du khi deux avec correction de Yates
 - 3.2. Test exact de Fisher
4. Existence d'un facteur de correction ?
 - 4.1. Le tableau 2 x 2 x 2
 - 4.2. Le test de Mantel-Haenszel
5. Bibliographie

Initiation à l'analyse des données

1. Introduction
2. Un exemple pour comprendre les notions en jeu dans les méthodes factorielles
 - 2.1. La notion de structure dans un tableau de données
 - 2.2. Exemple à 2 dimensions
 - 2.3. Exemple à 3 dimensions
 - 2.4. La notion de distance
3. Les méthodes factorielles d'analyse des données
 - 3.1. Différentes méthodes pour différents tableaux de données
 - 3.2. L'analyse en Composantes Principales (ACP)
 - 3.2.1. Objectifs
 - 3.2.2. Une transformation du tableau de données
 - 3.2.3. Analyse du nuage des individus
 - 3.2.4. Analyse du nuage des variables
 - 3.2.5. La dualité
 - 3.2.6. Aides à l'interprétation
 - 3.2.7. Exemple : caractéristiques pondérales des carcasses de bovins
 - 3.2.8. En résumé
4. L'analyse factorielle des correspondances
 - 4.1. Le tableau de contingence
 - 4.2. Transformation du tableau de données
 - 4.3. La ressemblance entre profils
 - 4.4. Construction des nuages et ajustement
 - 4.5. La dualité
 - 4.6. Interprétation d'une AFC
 - 4.7. L'inertie

5. Une méthode pour traiter les données d'enquête : l'Analyse Factorielle des Correspondances Multiples (AFCM)
 - 5.1. Les données
 - 5.2. Apurement et homogénéisation
 - 5.3. Transformation du tableau de données
 - 5.4. Construction des nuages et ajustement
 - 5.5. L'interprétation des résultats
 - 5.5.1. Le diagramme des valeurs propres
 - 5.5.2. Les aides numériques à l'interprétation
 - 5.5.3. Les représentations graphiques
 - 5.5.4. Description des facteurs
 - 5.5.5. Les éléments supplémentaires
6. Les méthodes automatiques de classification
 - 6.1. Principe
 - 6.2. La Classification Ascendante Hiérarchique (CAH)
 - 6.2.1. Quel tableau soumettre à l'analyse ?
 - 6.2.2. Le principe de l'algorithme de regroupement
 - 6.2.3. Choix du critère d'agrégation
 - 6.2.4. CAH à partir des coordonnées factorielles
 - 6.2.5. Le diagramme des indices de niveau
 - 6.2.6. L'arbre hiérarchique ou dendrogramme
 - 6.2.7. La description des classes
 - 6.2.8. La complémentarité analyse factorielle et classification
7. Pour aller plus loin en analyse des données
8. Bibliographie
9. Outils logiciels
 - 9.1. Liens internet
 - 9.2. L'aide en ligne des fonctions des librairies ade4 et mva
 - 9.3. Liste des fonctions de la librairie ade4
 - 9.4. Liste des fonctions de la librairie mva

Analyse de la variance

1. Introduction
 - 1.1. Le modèle linéaire et son cadre d'application
 - 1.2. Quelques mots de plus de vocabulaire
 - 1.3. Mise en œuvre pratique
2. Présentation et premiers regards sur les données
 - 2.1. Les données
 - 2.2. Lecture
 - 2.3. Quelques statistiques descriptives et graphiques
3. Analyse de variance à 1 facteur, en blocs
 - 3.1. Graphiques complémentaires sur la variable étudiée
 - 3.2. Analyse de la variance et graphiques associés
 - 3.3. Test complémentaire
4. Analyse de variance à 1 facteur, en blocs, avec problème d'hétérogénéité de variance
 - 4.1. Graphiques complémentaires sur la variable étudiée
 - 4.2. Analyse sur variable transformée
5. Analyse de variance à 2 facteurs croisés, en blocs.

- 5.1. Graphiques complémentaires sur la variable étudiée
- 5.2. Analyse de la variance
- 5.3. Test complémentaire
- 6. Bibliographie

Arbre de régression, arbre de classification

- 1. Introduction
 - 1.1. Le principe de base
 - 1.2. L'algorithme
- 2. Arbre de régression
 - 2.1. Présentation des données
 - 2.2. Lecture des données Rwanda sous R
 - 2.3. Calcul d'un arbre de régression
 - 2.4. Paramétrage de l'arbre et aide à l'interprétation